DOCUMENT RESUME

ED 422 650 EA 029 336

AUTHOR Kirst, Michael; Bomstedt, George; Stecher, Brian

TITLE A Plan for the Evaluation of California's Class Size Reduction Initiative.

INSTITUTION Policy Analysis for California Education, Berkeley, CA.; Rand Corp., Santa Monica, CA.; American Institutes for Research in the Behavioral Sciences, Palo Alto, CA.; WestEd, San Francisco, CA.

SPONS AGENCY California State Board of Education, Sacramento.
PUB DATE 1998-04-00
NOTE 35p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Diego, CA, April 1998).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Class Size; Educational Assessment; Elementary Secondary Education; Evaluation; *Evaluation Needs; Information Needs; Models; Program Evaluation; *Teacher Student Ratio
IDENTIFIERS *California; *Class Size Reduction

ABSTRACT

In July 1996, California began its Class Size Reduction (CSR) Initiative. To gauge the effectiveness of this initiative, an analysis of its objectives and an overview of proposed strategies for evaluating CSR are presented here. An outline of the major challenges that stand between CSR and its mission are provided. These include logistical challenges, financial challenges, the impact on teaching and learning, unanswered questions in the literature, and the value of a systematic evaluation of the program. The text outlines the design and key research questions that must be addressed by any evaluation plan and recommends six principles that call for a single, integrated evaluation; a comprehensive review; and a summative evaluation. A conceptual model of the evaluation plan is also provided. Overviews of how CSR will affect the following areas are included: state, district, and school policymaking; resource allocation; integration with other reforms; teacher quality, assignment, and training; classroom practices; parental involvement; and student outcomes. It is claimed that students' engagement with schooling, as measured by attendance, promotion/retention, homework completion, and frequency of disciplinary actions, may also change with the introduction of reduced class size and therefore should be assessed as part of the evaluation of student outcomes. A methodology section discusses evaluating the implementation of the class size reduction initiative with discussion on overall design, data collection, sampling, data analyses, an advisory board, and deliverables. Contains 29 references. An appendix discusses the importance of evaluating class size reduction to California educational policy. (RJM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 422 650

EA

A Plan for the Evaluation of California's Class Size Reduction Initiative

Submitted to the

California State Board of Education

By

Policy Analysis for California Education (PACE)

Rand Corporation

American Institutes for Research

West Ed

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
M. Kirst
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Presented to the annual meeting of the American Education Research Association by Michael W. Kirst, Stanford University, April 1998, San Diego

EA 029 336

BEST COPY AVAILABLE

Introduction

July of 1996 marked the beginning of California's Class Size Reduction (CSR) Initiative, a bold state effort to boost student achievement by limiting the size of kindergarten through third-grade classes. The program is a response to the continuing poor performance of California students. California ranks at the bottom on the National Assessment of Educational Progress (NAEP) fourth-grade reading achievement: In 1994 fewer than 20 percent of the state's fourth-graders scored at the "proficient level," and more than half did not even reach the "basic level," a benchmark that indicates only partial mastery of grade-level reading.

Prior to implementation of CSR, there were on average 30 students to every teacher in California public schools. This was the highest student/teacher ratio in the nation. The CSR program reduces this ratio to a target of 20 or fewer students per class.

The potential benefits of this dramatic class size reduction are considerable. Among these are:

- Teachers may be able to spend more time on teaching, and less time on managing student conduct, administrative work, or other non-instructional tasks.
- Students may receive more individual attention. Smaller classes may make it possible for teachers to monitor each student's work more closely and provide feedback and instruction that is better tailored to each student's needs and capacities.
- Educational reforms may have a better chance to succeed due to CSR. Smaller class sizes may make new approaches more tractable, which may in turn encourage teachers to implement changes authorized by the state or district, or to experiment with innovations of their own design.
- Parental participation in schools and schooling may increase. Smaller classes may give teachers more time to interact with parents, and parents may perceive a greater ability to contribute to their children's education.
- School climate may improve. With students receiving more personal attention, teachers feeling more efficacious and less bogged down by

extra-instructional demands, and parents more engaged in school functions, the overall level of morale at schools may be enhanced.

These positive outcomes may be mutually reinforcing. For example, as teachers spend more time on individualized instruction, the incidence of student misbehavior, tardiness or absenteeism may decrease, which may in turn encourage teachers to employ the types of innovations that require a higher degree of student cooperation and attention. Furthermore, many teachers who had previously spent more time keeping order in overcrowded classrooms than they did teaching, may now find it possible to give each child mastery over core literacy and analytic skills. Educational field trips and projects that were once impossible because of large class sizes now become possible. These new opportunities are likely to improve teacher morale and may serve as an antidote to the chronic problem of "burnout." Anecdotes of relieved and energized teachers set the tone of every discussion and news article about the initiative. For all these reasons, CSR has the potential to reverse years of decline and serve as a model for other states.

The initiative also represents a major financial commitment on the part of California to improving the quality of its public schools. CSR dwarfs other ongoing reforms in the state and across the nation. With a FY 97 price tag of over $1 billion, or $800 for every participating K-3 student, it represents by far the largest educational reform in the history of this, or any other, state.[1] Its impact is likely to be felt nationally since California currently educates more students than any other state, about one of every eight students in the nation. This school year 1.9 million young children will be assigned to smaller classes because of the initiative. Although participation is not mandatory, over 95 percent of California's districts took part, attesting to the popularity of the initiative. In a state with one of the lowest per-pupil expenditures in the country, many see the CSR's injection of new funds into early grades education as cause for celebration.

A Complex Undertaking

Although the promise of the CSR program is great, so are the challenges to implementing and maintaining it. Identifying these challenges, and monitoring their impact, may prove essential if CSR is to fulfill its considerable potential. An outline of major challenges that stand between CSR and its mission is therefore presented in order to help focus evaluation plans.

[1] Legislators enacted CSR as part of a package of reforms to raise reading achievement. CSR funds were accompanied by $80 per student for K-3 reading materials.

Logistical Challenges

Working through the logistics of the initiative's space and personnel demands has been a sizable task. The urgent need it has created for new teachers and more classroom space has refocused much of the state's education agenda around the CSR initiative. For example, in 1996, the first year of the initiative, districts hired 18,000 new teachers to cover the new classes it generated. Almost one quarter of these teachers were uncredentialed, placing new demands on alternative certification programs. In the next two years the demand for new teachers will be greater and the proportion of under-prepared teachers will increase.[2] Teachers in these smaller classes—both new and experienced—will need to learn to teach small groups effectively for improved student achievement. Additional staffing problems are occurring in other grades and program areas because some of the teachers from upper grades and from programs addressing special-needs students have been attracted to CSR classrooms. This reassignment may cause less qualified and less experienced teachers to be assigned to the more challenging and larger classes.

To be successful, the program will require administrative support; but the time and energy the program absorbs from the state's educators outside the classroom, including principals, superintendents, their staffs, as well as administrators at the State Department of Education, is time unavailable for other functions. Similarly, while CSR enhances education opportunities for students in the primary grades, it may pull resources away from the higher grades. Finally, the initiative may conflict with and divert resources from other state and district reform programs, thereby interfering with ongoing efforts to improve schools. For these reasons, any evaluation plan must consider how CSR, and the recourses that it demands, affects the overall network of initiatives now in place to improve California's educational standards.

Financial Challenges

The initiative provides uniform funding per additional reduced size classroom to all districts, irrespective of local costs. This funding method may accentuate inequities in educational resources and services across the state. For example, according to the California Research Bureau, the cost in FY 97 to districts for

[2] In most participating schools, Year 1 implementation reduced class size in one or two grades. Schools were required to reduce class size in first-grade classrooms first. Second grade was second priority, and either third grade or kindergarten was third priority. In 1997, some schools will implement the initiative in three or even four grades, from kindergarten through third grade.

reducing class size ranged from zero to over $1,000 per student. In most districts the FY 97 state allocation of $650 per student was below cost. While the enhanced FY 98 state appropriation will allocate more than enough to cover costs for many districts, others will still need to redirect funds from other budget lines to cover the additional cost of implementing the CSR program.

Furthermore, in order to participate, schools must find additional classroom space. Early data suggest that the larger and more urban districts, as well as those faced with the highest enrollment growths, may have already reached space limits. These schools may be able to implement the program in just one or two grades. Many have had to convert libraries, computer rooms, and music rooms into classrooms. The majority of school districts are using portable classrooms to implement CSR, but many sites do not have places to put such structures. In some schools, CSR teachers "share" classrooms; the children are actually assigned to classes of 40 with two teachers. In addition, with the many new teacher openings and a statewide shortage of experienced credentialed staff, veteran teachers are transferring from lower- to higher-paying districts, possibly exacerbating resource differentials between the poor and the rich.

Impact on Teaching and Learning

CSR may provide important opportunities for improvements in teaching and learning. However, unless teachers make conscious changes in their teaching strategies, the smaller classes alone may not improve students' performance. Teachers and principals need continuous feedback about strategies that work to help them make the initiative succeed.

CSR comes at a time when California enrollments are growing and the proportion of children who are not proficient in English is at an all-time high—one out of three in grades K-3. In addition, one out of four children attending California's public schools lives in poverty; the same proportion live in single-parent homes. Ultimately, smaller class sizes may help effect transitions of non-native speakers and other special needs children into the curriculum. In fact, research indicates that such students are among those who benefit the most from smaller class sizes. However, current conditions are such that it will be a great challenge for CSR to meet the language instruction and other special needs of all the state's schoolchildren.

Unanswered Questions in the Literature

The balance of evidence suggests that substantial reductions in class size do improve student achievement (Blatchford and Mortimore, 1994; Finn and Voelkl, 1992; Glass et al., 1982; Illig, 1997; Mosteller, 1995). The effects are strongest for students in the early primary grades (Educational Research Service, 1980), for low-achieving students (Angrist and Lavy, 1997; Krueger, 1997), and for students from poor socio-economic backgrounds (Finn and Achilles, 1990). It also appears that achievement gains are greater when classes are smaller—20 students or fewer (Glass and Smith; 1978). Reducing class size also appears to decrease retention and referrals to special education (Snow, 1993; Illig, 1997), and it boosts teachers' morale and job satisfaction (Glass, et al., 1982; Shapson, et al., 1980).

However, "the question of why these effects are realized remains largely unanswered" (Finn and Achilles, 1990). As a result, research offers little guidance for implementing the reform on a large scale. What should teachers do to take advantage of small classes? What specific professional development should districts provide to newly-hired staff? How well does it meets the needs of limited English proficient (LEP) students; and whether it represents the best use of educational resources.

The few studies of teaching practices in reduced-size classes do not provide clear answers to these questions. Cahen et al. (1983) studied four classes intensively after enrollments were reduced and found that changes in practice did occur, but were not dramatic. "Teachers and students were happier and more productive" but "the process of instruction looked very much the same" (p. 201). Shapson et al. (1980) found similar results in a study of fourth-grade classes in Toronto. These researchers noted marked improvements in teachers' attitudes, but little corroborating evidence of changes, for example, in the proportion of time teachers allocated to whole class, group, or individual activities.

The literature is silent on a number of other important questions. How should administrators allocate scarce resources, such as classroom space and experienced teachers? The Tennessee STAR program, which was the largest and best-controlled study of class size reduction, involved 79 schools (Mosteller, 1995). In general, the schools all had the necessary facilities and were able to recruit trained staff to make the program operate smoothly. Other experiments were conducted on smaller scales and did not have to address such problems. Consequently, the literature does not address the larger policy questions that arise when implementing such a reform statewide (Murnane and Levy, 1996; Mitchell and Beach, 1990). For example, what supportive policies are needed to realize the benefits of class size reduction? How does class size reduction compare to other

uses of educational resources? These are questions which should be explored in order to monitor and advance class size reduction in California. A thorough and independent evaluation of this expansive effort is therefore needed in order to obtain timely responses to these vital policy issues.

The Value of an Evaluation

The following section outlines an evaluation plan that addresses the implementation of the program as well as its effects on schools, classroom practices, and student achievement. The evaluation should be both *formative*, providing feedback for improvement during the life of the program, and *summative*, generating results on cumulative impact.

The formative component provides state education policymakers with information about how best to implement this program. It identifies problems as they arise, points to potential solutions and provides insights into how educators throughout the state are responding to CSR issues.

The summative component helps answer the question uppermost in people's minds: *Do smaller classes help improve achievement?*

In addition, however, the summative component should uncover the program's particular effects:

- *Has the reform led to beneficial changes in classroom practices?*
- *Under what circumstances has the reform been most successful?*
- *What effect, positive or negative, has the reform had in addressing inequities among California students?*
- *How have other educational programs and upper grades been affected by the focus given to the early grades?*
- *How have limited English proficient (LEP), special education, and other "at-risk" students been affected?*

Finally, the summative component should help answer those questions that must be addressed for all major public initiatives—whether it attains its goals, whether it has unanticipated consequences that need addressing, and whether it is worth the energy and resources it commands.

The findings from an evaluation of the CSR Initiative will have relevance for the current national discussion about what works to improve student achievement, which is the goal of numerous ongoing state and national reform efforts. At the moment, a growing list of states (including Connecticut, Florida, Georgia, Hawaii, Iowa, Kansas, Louisiana, Massachusetts, Michigan, Minnesota, Nevada, New York, and Utah) are implementing or considering reductions in class size.

The Design and Key Research Questions of the Evaluation Plan

The evaluation design recommended is guided by six principles that emerged from conversations with state-level policymakers, superintendents, principals, teachers, and representatives of research organizations and professional groups.

- A single, integrated evaluation is preferable to a set of small-scale studies on topics of concern. Assessing CSR through disconnected projects would probably fail to provide a meaningful picture of the whole initiative or capture the relationships among the multiple actions needed to implement the program.
- CSR is a system-wide intervention that may affect student performance, teacher practices, and district expenditures and policy. For this reason it is essential that the study be comprehensive, addressing all relevant issues. A comprehensive review will make it possible to determine under what conditions and for what reasons CSR has an impact.
- The evaluation should provide information to improve implementation as well as to determine whether the initiative ultimately succeeds. The initial "formative phase" evaluation will produce information to aid ongoing decision-making at all levels of the system. State leaders and educators emphasized the value of ongoing feedback and information about the status of CSR implementation, problems encountered at all levels, their resolutions, and innovative practices associated with class size reduction.
- The summative evaluation should answer questions about the relationship of reduced size classes to student achievement and to educational practices throughout the system.
- A longitudinal approach to the evaluation is essential because CSR will take years to be fully implemented. An immediate snapshot study will not adequately reflect the changes that occur as the program matures, while a summative approach alone would not reveal mid-course corrections that may be needed.

- The evaluation needs to be rigorous and objective so the findings will be credible to both supporters and skeptics.

Conceptual Model of the Evaluation Plan

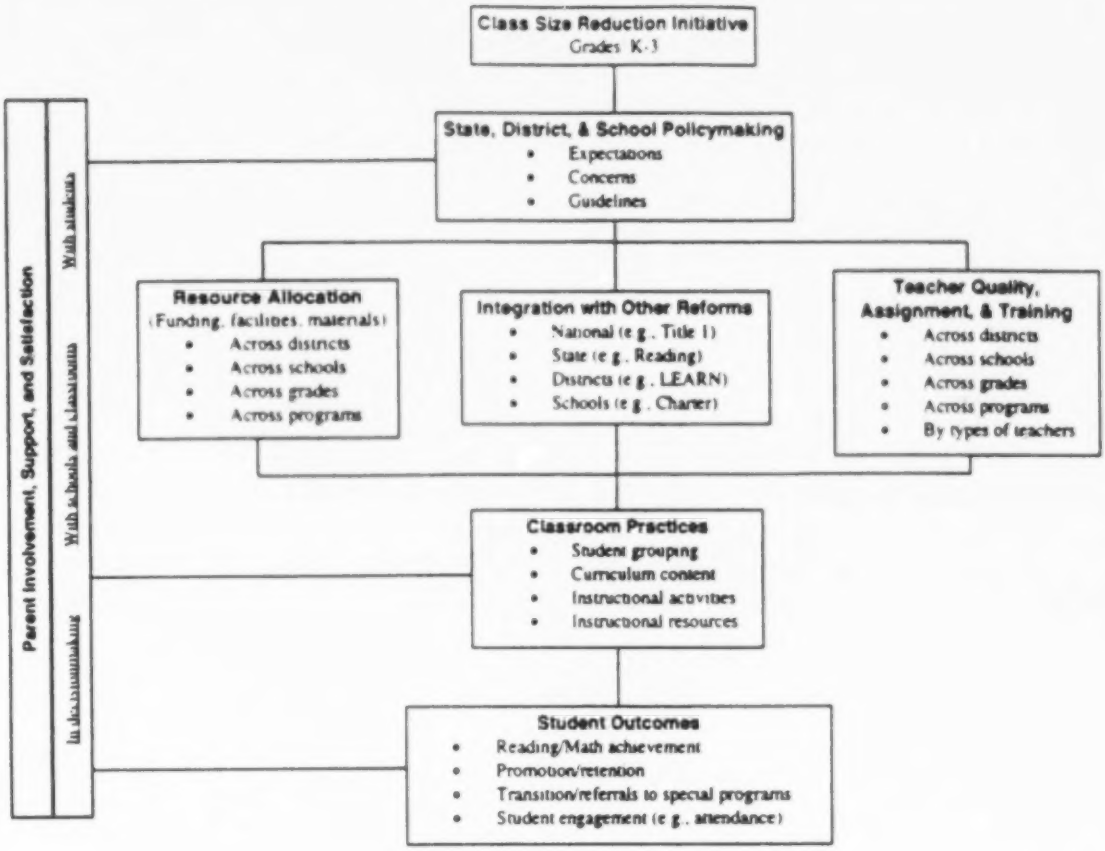
The underlying conceptual model that should guide the evaluation is shown in figure 1. It has built into it the connections among the seven major issues identified above. The model begins with an examination of how district and school policies might have been affected by the State's CSR initiative and how these policies relate to resource allocation, other ongoing reforms, parental involvement and support for the program, and to teacher quality and training. These factors are assumed to relate to classroom practices, which in turn are assumed to relate to student outcomes.

Each of the various components in the model gives rise to a set of evaluation questions that will be elaborated in the remainder of the section. However, the overall guiding questions are simple:

- *How does the reform effect students?*
- *What factors account for its success or failure?*
- *What is the relationship between the program's benefits and costs?*
- *How can the reform be implemented most effectively?*

We now turn to a discussion of each of the seven components and the research questions that these components should address.

Figure 1—The Impact of the California Class Size Reduction Initiative



State, District, and School Policymaking

Participation in the class size reduction program is voluntary. Districts have choices about the number of grade levels and schools that will reduce class size to 20, and these decisions can change annually. A number of factors influence these decisions, with the local context playing as great a role as the state guidelines. Policymaking at the state, district, and school levels will be considered to understand how beliefs, concerns, and context influence decisions regarding class size reduction. For example, the match between state, district, and teacher expectations for the goals of CSR may influence its success. The degree to which existing policies and practices are aligned with CSR is also likely to have an impact on the initiative's success. Consequently, how CSR implementation is affected when local policies (e.g., school and district literacy initiatives, professional development activities, and standards and assessments) complement

BEST COPY AVAILABLE

the goals of CSR, should be investigated. Collective bargaining agreements between teachers and school systems are another part of the local policy context that should be examined.

Key research questions regarding Policymaking should include:

- What goals do state policymakers and district and school administrators have for CSR, and what are their concerns regarding implementation? How do differences in expectations and/or concerns affect their actions?
- Do policymakers share a common set of expectations about how CSR can influence student learning? Are these expectations different from those of teachers?
- Which educational policies, regulations, and labor agreements facilitate or impede the effective implementation of CSR?
- How are implementation decisions made about which grade levels or classrooms participate in CSR each year? How does participation in these decisions vary across districts, and why?

Resource Allocation

Preliminary studies of FY 97 CSR implementation have raised a number of questions about the reallocation of resources within and among schools and districts (Office of the Legislative Analyst, 1996; Blattner et al., 1997). In the first year, the average cost per student of CSR exceeded the state allocation of $650 per pupil by 21 percent, or $140 per student. Research suggests that this funding shortfall may have affected the relative distribution of funding between elementary and secondary schools as well as the distribution of resources within schools between the targeted early primary grades and subsequent grades. FY 98 funding increases alleviated this shortfall, on average. However, some districts are still experiencing surpluses and others, deficits, through this program. It is important that the evaluation examine the impact of implementing this program in these two types of districts. What has been the effect on the allocation of space and facilities in districts and what other programs and/or educational activities have been affected by CSR classroom demands?

Key research questions regarding Resource Allocation should include:

- What is the impact of the CSR program on equality of funding for education across districts? Across schools? Across subpopulations of students such as LEP students, minority and low-income students, and those with special needs?
- How does CSR affect district revenues and expenditures? Within districts, how does the initiative affect school spending levels on operations and facilities?
- Within schools, how does CSR affect resource allocation across grades, instructional support services (e.g., libraries, media centers, counseling) and instructional programs (e.g., art, P.E., music)?
- How does CSR affect the distribution of resources—funding, space, and materials— across primary and secondary schools and across programs within districts? What strategies have schools used to make space for new classrooms? What tradeoffs were made and why? How permanent are they?

Integration with Other Reforms

The CSR initiative is recent, large, and potentially of great significance, but it is not the only reform initiative involving California schools. Other ongoing reform efforts include state initiatives on literacy, professional development, and charter schools; district initiatives such as LEARN in Los Angeles; privately funded initiatives like the Annenberg Challenge, which supports the Bay Area School Reform Collaborative (BASRC) in the San Francisco area and the Los Angeles Metropolitan Project (LAMP), Accelerated Schools, and the Coalition of Essential Schools; as well as a variety of local technology, science, language proficiency, and arts efforts.

CSR has the potential to enhance these existing reform initiatives or to detract from them. On the one hand, a reform initiative which integrates school literacy and professional development can be focused on how literacy may be improved by new small class instruction practices. However, at the same time class size reduction may distract schools from the pursuit of other school reform initiatives. For example, some elementary science labs have been closed to provide additional classroom space for CSR. Similarly, initiatives to strengthen the middle school curriculum may lose momentum because of the concentrated effort needed to implement CSR. Schools undergoing restructuring may find that the attention focused on class size reduction displaces efforts to bring about systemic change.

Key research questions regarding Integration with Other Resources should include:

- How does CSR interact with large categorical programs like special education and Title I? How does CSR affect programs serving limited English proficient students and special education students?
- Do CSR implementation approaches differ by district/school characteristics? For example, do low-revenue districts implement CSR differently than high-revenue districts?
- Is CSR integrated with the district's master planning efforts, or does it exist independently of other initiatives the district/school may wish to pursue? Does CSR serve as a catalyst to enhance coordination of existing reform efforts, or does it provide a diversion from more systematic efforts?
- What is the nature of other reform efforts in the district/school at the time CSR is introduced? How are other school reform efforts affected by the introduction of CSR? Are resources (dollars, time, and people) redirected from other reform efforts to assist in CSR implementation? Does CSR affect changes in staff assignments at the district or school levels?

Teacher Quality, Assignment, and Training

Human resources are an essential part of CSR. The effectiveness of the reform will depend in large measure on the quality of the teachers in the system, choices about which teachers are assigned to smaller classes, and the preparation teachers receive for these classes. There are reasons to be concerned on all fronts. First, the demand in the first year of the program alone vastly exceeded the supply. The Legislative Analyst's Office (1997) reports that 30 percent of the new hires were not credentialed, 24 percent were granted emergency permits, and 6 percent were enrolled in university programs but not credentialed. Only 14 percent of the new hires had more than five years of teaching experience. Overall, it appears as though underqualified teachers may be entering California in large numbers.

The evaluation should track the assignment of these new teachers, as well as of those already in the system. As noted previously, some districts report that teachers at higher grade levels and special education teachers are requesting transfers to smaller classrooms, leaving less well prepared teachers to tackle these demanding assignments. There have also been reports of highly qualified teachers transferring from urban districts to suburban districts as new positions are created. Therefore it is important that the evaluation also investigate professional

development programs used to train teachers and to examine whether the support being provided is adequate to ensure that effective teaching is occurring in small classes. Finally, it is important that the evaluation examine whether and how CSR has affected teachers' attitudes toward their job and their engagement in it.

Key research questions regarding Teacher Quality Assignment and Training should include:

- How is CSR affecting the recruitment and assignment of teachers across districts, schools, grades, and special programs? How does collective bargaining influence this process?
- What are the qualifications and experience of teachers assigned to smaller classes? What is happening to the qualifications of teachers in classrooms with high concentrations of limited English proficient students, minority students, and students with special needs?
- What professional development activities and support are provided for teachers assigned to smaller classrooms? How do these activities differ across categories of teachers (e.g., noncredentialed, newly credentialed, and experienced but new to primary grades)? What type of training do teachers assigned to smaller classes receive with regard to language instruction strategies for limited English proficient students?
- How does CSR affect teacher satisfaction and attitudes toward teaching and students? How do the attitudes of teachers in smaller classes affect students' learning opportunities and potential?

Classroom Practices

Little is known about which classroom practices are most effective when class size is reduced. Some advocates of smaller classes argue that reducing class size enables teachers to have more individual contact with students. Smaller classes also reduce teachers' burdens associated with discipline, paperwork, and other noninstructional duties and free them to devote more class time to teaching. From this perspective, the key advantage of class size reduction is that it permits *more* activities and interactions to occur—more contact, more feedback, and more exposure to curriculum.

Other proponents argue that the principal advantage of class size reduction is that it permits *different* activities and interactions to occur. According to this argument, the reduction in noninstructional demands coupled with a better knowledge of students' individual needs allows teachers to engage in different

kinds of interactions, including student-centered learning in which the teacher acts as facilitator rather than dispenser of knowledge, extended project-based learning, increased emphasis on higher-order skills such as problem solving, and richer literacy experiences. In this view, class size reduction permits changes in the nature of teacher-student interactions and in the content of the curriculum.

It is important that the evaluation of CSR examine whether and how it has affected the rate and/or the nature of the activities and interactions teachers have with the children in their classrooms.

Key research questions regarding Classroom Practices should include:

- What changes have occurred and are occurring in teaching practices as a result of CSR, including changes in emphasis or coverage of different topics, methods of instruction, and the range of learning experiences?
- What types of language instruction strategies/models are used in CSR classrooms for limited English proficient children in CSR classrooms? Do changes in instructional practices differ across districts, classrooms, and categories of students (e.g., LEP, minority, and special education students)?
- What changes occur in the availability and allocation of instructional support personnel and other resources (e.g., staff development, curriculum guidance, and opportunity for collaboration with other teachers)?

Parental Involvement

Many educators believe that parent involvement will improve children's educational success. Certainly, the available research supports the belief that parents matter. At the elementary school level, research has demonstrated an association between parent involvement and fewer behavioral problems (Comer, 1984), lower dropout rates (NCES, 1992), higher student achievement (Muller, 1993; Stevenson and Baker, 1987; Reynolds, 1992; Kohl, 1994; Klimes-Dougan et al., 1992), and children's perceived level of competence (Wagner and Phillips, 1992).

Some proponents expect CSR to have a positive effect on parent involvement, although the mechanisms that would facilitate such an effect are far from clear. It might be that those parents whose children are in smaller classes view the district as being more concerned about children and about their child. Hence, they may

feel their support is less essential. On the other hand, parents may feel less intimidated about "bothering" a teacher who has fewer students, believing that she will have more time for them and their concerns. In fact, teachers may actually have more time for parents, and may more actively seek their involvement at both the classroom and school levels.

The evaluation of CSR should examine whether teachers have more time for parents and whether parents are in fact spending more time with the teachers and participating more in the education of their children. It should also examine parents' satisfaction with the school and whether their attitudes about the quality of their children's have changed.

Key research questions regarding Parental Involvement should include:

- To what extent have parents been involved in decisions about grade participation, reallocation of resources and space, and the assignment of students to classrooms at the district or school level?
- Do parents feel more welcomed in their children's classrooms? Do they have more parent-teacher conferences? Do class assignments and activities receive increased or decreased amounts of parent participation?
- Do CSR parents believe that their children are receiving a better education (e.g., more individualized attention)? How does the initiative affect their behavior toward their children? Does CSR affect parents' satisfaction with the teacher, the school, or the district?
- Has the range or intensity of parent involvement programs and efforts at the school declined or increased as a result of CSR? Has the amount and nature of parent involvement in the schools changed as a result of CSR?

Student Outcomes

The primary motivation for reducing class size is to improve student learning. The main criterion for academic achievement in the evaluation should be performance on the new standardized reading and mathematics tests (STAR) recently adopted by the State. However, because many standardized tests measure mainly comprehension, it is important that the evaluation examine students' oral reading ability as well.

For English language learners whose primary language is Spanish, Spanish versions of standardized tests adopted by selected districts should be used. In

addition, it is recommended that the evaluation measure Limited English Proficient students' reading readiness. Standardized reading tests may not be sensitive enough to capture gains in English language development that have occurred with the introduction of reduced class size. More specifically, while some students may not be able to read per se, they may have developed language skills, such as word recognition, that would show up on a special reading readiness assessment.

Students' engagement with schooling, as measured by attendance, promotion/ retention, homework completion, and frequency of disciplinary actions may also change with the introduction of reduced class size and therefore should be assessed as part of the evaluation of student outcomes. Changes in referrals and transition rates of students into and out of special education, bilingual education, or other programs should also be measured, along with teachers' views about long-term improvements in students' readiness for a new grade.

Key research questions regarding Student Outcomes should include:

- Has student achievement in reading and math improved since CSR began? Have promotion rates to the next grade changed as a function of CSR? Do next grade teachers perceive improvements in students' preparation to master grade-level material?
- Has there been an increase or decrease in transition rates into or out of special programs (e.g., sheltered English programs, resource classes, reading interventions)?
- Are students more engaged in school (in terms of attendance, behavior, and homework completion)?
- Has reading readiness improved for ESL students with the introduction of CSR?
- Do any of the relationships between class size and student outcomes vary on the basis of school, teacher, classroom practices, and/or student characteristics (e.g., do limited English proficient students benefit more than English proficient students)?
- Are changes in classroom practices, due to CSR, associated with changes in students' educational outcomes?

Methodology

Overall Design

Because full implementation of class size reduction will take several years,[3] the evaluation should collect data annually for three years beginning in the current 1997-98 school year. Archival data should also be used to establish pre-CSR baselines against which to measure change. For some topics (e.g., classroom practices), however, establishing such a baseline will not be possible. In these cases, the status at the beginning of the study and changes thereafter should be described. However, for those classrooms and schools where CSR was not fully implemented by the beginning of the current school year (i.e., 1997-98) baseline data can and should be collected.

In order to link and aggregate the information gathered at different levels of the system, a nested sampling design of districts, schools, and classrooms should be implemented. Stratified random samples should be taken, respectively, of districts representing the state as a whole, sample schools within selected districts and, finally, teachers and classroom within selected schools. Particular attention should be paid to the relationship between CSR and student outcomes, including achievement. To this end, achievement tests of successive cohorts of fourth-grade students using a time series (post-test only) design should be used.[4] CSR does not apply to the fourth grade, and for that reason systematic changes in fourth-grade performance can serve as an outcome index with which to measure the effects that CSR has on learning. Succeeding fourth-grade samples will represent cohorts who have had increasing exposure to reduced size classes, permitting analysis of both overall CSR "dosage" as well as the grade levels at which exposure to CSR first occurred.

The sampling plan adopted should have sufficient statistical power to detect substantively important mean differences for successive fourth-grade achievement outcomes, overall as well as for subgroups based on gender, income, urban/rural residence, and LEP status. Other analyses would complement this approach, i.e., the sampling plan adopted should have sufficient statistical power to detect substantively important percentage differences when analyzing the data drawn

---

[3] The class size reduction program began in 1996-97, and its greatest expansion will occur during the school years 1996-97, 1997-98, and 1998-99. Most schools in California had reduced-size classes in one or two grade levels in 1996-97. Most will have smaller classes in two or three grade levels in 1997-98, and in three or four grade levels in 1998-99.

[4] Data from a new state testing program which will use the same commercial test in all schools throughout the state and is scheduled for use in the spring of 1998.

from the surveys described below, including differences for subgroups based on income, urban/rural residence, and LEP status.

Data Collection

It is important to have a data collection plan that is the most appropriate for the questions listed in the preceding section, while at the same time minimizing the burden on respondents. Table 1 shows the data collection plan the for the evaluation of CSR.

Table 1—Data Collection Methods by Research Topics

	Topic Area					
Methods	Resource Allocation	Other Reform	Teacher Quality	Classroom Practice	Parent Involvement	Student Outcomes
Existing Databases						
CBEDS	X					
SDE District Financial Data	X					
Cost-of-Education Index	X					
Administrative Records	X	X				X
Standardized State Tests						X
Mail Surveys						
District Administrator	X	X	X		X	
School Principal	X	X	X	X	X	X
Classroom Teacher		X	X	X	X	X
Parent	X		X	X	X	X
Case Studies						
Districts	X	X	X		X	
Schools	X	X	X	X	X	X
Classrooms	X	X		X	X	X

Existing Databases. Several existing state or district databases should be used to address questions relating to resource allocation and teacher preparation, including (1) the California Basic Educational Data System (CBEDS), which includes information about teachers' backgrounds and personnel assignments, and district and school information, (2) district financial records maintained by the State Department of Education (SDE), and (3) the Cost-of-Education Index (CEI), which permits adjustments for geographical differences in costs.

Standardized Tests. The plan is to analyze data from the standardized, statewide achievement test, which is to begin in the spring of 1998, and which presumably will also be available on CBEDS. Data from these tests should be complemented

with district and school records to monitor changes in other student outcomes, including attendance, retention/promotion in grade, transition rates into and out of special programs, and behavioral problems.

An oral reading test such as the NAEP Integrated Reading Performance Record (Pinnell *et. al.* 1995) should also be administered to successive samples of fourth-grade students beginning in the spring of 1998 through the spring of 2000 to ensure that more than reading proficiency is being assessed.

All students will be required to take the state's new STAR test in English, except for students with Individual Education Plans (IEPs) which explicitly exempt them from testing requirements. As indicated above in discussing reading achievement, it may not be sensitive enough to measure achievement in those limited English proficient students who have not been taught solely in English. Hence, data from tests taken in languages other than English should also be analyzed as part of the evaluation plan.[5] In this regard, many districts opt to administer tests in Spanish to Spanish-speaking students.

Finally, a reading readiness test (e.g., Woodcock-Johnson) should be administered as part of the evaluation to successive samples of fourth-grade LEP students beginning in the spring of 1998 and continuing through the spring of 2000 in order to better assess pre-reading achievement gains for LEP students. These are gains that might not be detected by the new standardized reading test.

Mail Surveys. The evaluation plan should include four survey instruments—one each to district administrators, school administrators, teachers, and parents.

District Survey: The district survey should be sent to superintendents, but it should be organized by district function—finances, personnel, instruction, facilities—so it can be delegated to staff with those particular administrative responsibilities. It should focus on questions about the use of facilities, changes in internal resource allocations, recruitment and hiring, teacher assignment practices, staff development, integration of CSR with the district's planning efforts, parent involvement, and opportunities forgone as a result of class size reduction.

School Survey: This should be sent to principals and should include questions about classroom organization (resource allocation), the use of school facilities, changes in teacher assignments (staff development), support for new staff, the

[5] The state is mandating such tests for students who have been enrolled in U.S. schools for less than one year and is encouraging such tests for all LEP students.

provision of equipment and materials, integration with other reform efforts and programs, and the involvement of parents in school activities.

Teacher Survey: This should focus on the effect of reforms on teaching and learning, contacts with parents, and selected instructional issues. It should also address teachers' professional development opportunities, curriculum coverage, classroom organization, and access to materials.

Parent Survey: This should address parents' participation in policymaking, contact with their children's school, support for learning, and satisfaction with class size reduction.

Case Studies, Observations, and Videotaping.

Case Studies: Case studies should be conducted in a limited number of districts, schools within those districts, and classrooms within those schools, to collect qualitative and process information that can only be obtained through open-ended interviews and through field observations and videotaping.

As appropriate to the entity studied (i.e., the district, school, or classroom), open-ended interviews should be conducted with administrators, principals, teachers, special program directors, school board members, union and parent representatives, as well as others. These interviews should cover issues related to district and school policies concerning class size reduction, including implementation (e.g., which grades to reduce first, facilities requirements, teacher hiring and assignment, allocation of resources), instructional support (e.g., staff development, teacher planning or collaboration, curriculum development), integration with other programs, and problems encountered and solutions adopted.

Classroom case studies should be the primary source of information for aspects of curriculum and instructional practices that are difficult to measure using surveys, including teachers' approaches to curriculum topics, their expectations, and the use of reform-oriented instructional strategies. Specific data collection strategies should include teacher and principal interviews, teacher logs, classroom artifacts, observations, and the videotaping of actual classroom instruction.

Interviews: Interviews with teachers should cover such areas as teaching background (e.g., years of experience, range of class sizes taught, credentials), curriculum and teaching practices, beliefs about the effects of class size reduction, perceptions of how practice changed with class size reduction (or for teachers in larger classrooms, perceptions of how practice might change), type and quality of instructional supports, type and frequency of contact with parents, and perceptions of student outcomes associated with smaller classes.

Logs and Artifacts: The collection of logs and classroom artifacts (annotated assignments and examples of student work) should focus on the content of the curriculum and the embodiment of instructional goals. A sample of teachers should also be asked to keep a daily log that collects information on curriculum topic coverage and emphasis on teaching practices, student activities, grading and homework, texts, and equipment. Teachers should keep these logs one week out of every four according to a schedule to be developed at the beginning of the study. These logs would be supplemented with copies of curriculum materials, including classroom and homework assignments, assessments, and samples of student work from a random sample of five students per class. Teachers should be paid for participating in the study and carrying out the extensive data collection activities that participation entails.

Observations and videotapes: Observations and videotapes should be used to examine instructional interactions and teacher practice variables that cannot be captured in interviews or inferred from artifacts. During the first year, a mathematics and a language arts lesson from each teacher should be videotaped, similar lessons being videotaped during the second year. This scheme is designed to balance the advantages and disadvantages of the two methods.[6]

Sampling

Mail surveys and case studies in a nested sample of districts, schools within these districts, and classrooms within these schools should be conducted as follows:

Mail Surveys. First, a stratified, random sample of districts should be selected with probability proportional to size, i.e., the number of fourth-grade classes each contains. In addition to size, stratifying variables should include median income, urbanicity (i.e., urban, suburban, rural), and share of students in the district with limited English proficiency. This approach assures that the largest school districts in the state would be in the sample, and that the state's diversity of settings and students is adequately represented in the sample.

Second, within the districts selected above, a random sample of schools that contain grades K-4 should be selected, an average of three schools per district, with a minimum of one school per district. The selection process should be the same as above, i.e., the probability of selection of classrooms in any one district should be proportional to the number of fourth-grade classes in the district after stratifying for share of limited English proficiency in the schools.

[6] The Third International Mathematics and Science Study (TIMSS) has demonstrated the value of videotapes for analyzing and comparing classroom practices (Stigler and Fernandez, 1995).

Finally, a sample of teachers and parents should be selected from the universe of teachers and parents in the schools selected above. Parents should be selected so as not to take only the active or engaged parents.

The sampling strategy utilized must provide samples large enough to reliably generalize the findings from the test sample to the state as a whole. Sampling should accommodate all fourth-graders as well as all major subgroups based on gender, region, urbanicity, ethnicity and percent of LEP students in the school. A survey completion rate of 80 percent for districts and schools, and 75 percent for teachers and parents, is assumed for the plan.

Case Studies. The districts, schools, and classrooms in which the detailed implementation and qualitative case studies would be conducted should be purposively selected from the sample of districts and schools as selected above. A nested design should also be used for this part of the project using classrooms generated as part of the overall sampling plan outlined above. The diversity of the state with respect to racial/ethnic composition should be represented in the case studies. At least two of the largest districts in the state, and at least two suburban and one rural district should be included.

Student Achievement Tests. Since the state achievement tests will be administered to all fourth-grade students throughout the state, the plan calls for data for all classrooms in the state to be used (see *Data Analyses*, below). All schools in the sampled districts that used the adopted state test in 1996-97, 1997-98, and 1998-99 in order to conduct analyses of student achievement *within* cohorts should also be selected (See Data Analysis, below). Based on current test use data in California, it is estimated that this subsample would include approximately 50 schools.

Data should also be collected in four districts where a Spanish language test, such as SABE/2, is given. These four districts will be selected from the pool of districts included as part of the overall sampling plan. Within those districts, two schools will be selected based on availability of Spanish language test scores, size, region, and a high percentage of limited English proficient students whose primary language is Spanish.

Finally, for a subsample of no fewer than 200 schools chosen from the schools included in the overall sampling design, no fewer than five students should be randomly selected to take an oral reading test in order to validate that high scores on the state's new standardized reading achievement test correlate highly with actual reading ability. A second sample of no fewer than five students in each of

the schools should be selected from their LEP and IEP populations to take an individual-level reading readiness test.

Data Analyses

Although the evaluation plan calls for the collection of different types of data (quantitative and qualitative), from a range of respondents (administrators, teachers, and parents) using a variety of methods (surveys, interviews, case studies), relating to a number of issues from resource allocation and classroom practices to student outcomes. Considerable effort should go into data preparation. In the case of quantitative data (such as state databases, surveys, fixed-choice interviews, and student outcome data), steps include data entry and verification, data reduction/simplification, the linking of comparable data from different sources, and the preparation of files for analysis. Similar functions would be performed for qualitative data, such as open-ended interviews, classroom artifacts, and logs, according to procedures described below.

Quantitative Data Analyses. Data analysis should provide valid and reliable results. The statewide distribution of information about each of the research questions should be examined, as well as differences in CSR effect variables between particular types of districts, schools, classrooms, and students. Relationships among the seven research issues should also be explored, providing associations between the research topics such as teacher experience and classroom practices, and resource allocation and parent involvement. More complex relationships should also be examined, including questions about the relative contributions of multiple factors. In addition, changes over time should be examined by repeating much of the data collection on an annual basis. The approach to longitudinal analyses should be similar to the cross-sectional analyses just described.

To determine whether class size reduction is related to an increase in reading and mathematics scores, average achievement scores by schools on California's new standardized test for successive fourth-grade cohorts in the springs of 1998, 1999, and 2000 should be examined. This analysis should include all fourth-grade students in California elementary schools who take the test. If CSR is having an impact on reading and mathematics achievement, a statistically significant improvement in scores in 1999 (compared to 1998) and again in 2000 (compared to 1999 and 1998) would be expected.

Analyses should then be refined by including teacher characteristics (e.g., years in the classroom and teaching credentials) by school and by grade level in the model to determine whether they have a significant relationship with reading and

mathematics achievement above and beyond their relationship with CSR measures. Finally, it is important to see whether any observed relationships between CSR and achievement hold for all regions, genders, racial and ethnic groups, and social classes. Therefore, all the analyses above should be repeated after dividing the sample using these variables.

One issue that needs to be addressed in the analysis of the successive fourth-grade cohort study is the introduction of a new test. As teachers and students become familiar with the test, an apparent achievement gain is likely. However, this observed "gain" may be the result of their growing knowledge of the format of the test or to "teaching to the test," rather than being due to CSR. Although test scores may improve because of teaching to the test, research shows that this gain cannot be replicated when the same content area is measured by a second test (Linn and Kiplinger, 1995). Therefore, in addition to the analysis of outcome data for successive fourth-grade cohorts, data from a sample of schools that have used the STAR test for the years 1995-96, 1996-97 (i.e., before it was chosen as the STAR test) should be analyzed to estimate how much gain might be expected from teaching to the test (i.e., in the absence of CSR). These results should be noted when reporting results relating the STAR to CSR assuming the sample of districts using the standardized test prior to its adoption are representative of the population of districts in the state.

The relationship between CSR and oral reading ability for a subsample of fourth-grade students should also be examined, using the same methods outlined above. Furthermore, the results should be compared with those generated by the 1994 NAEP oral reading assessment.

Given the large limited English proficiency (LEP) population in California, an important question is whether the relationship between CSR and achievement is the same for all students. Further, resources that might otherwise be used for special populations may now be used for CSR. Therefore, all of the analyses described above should be run separately as a function of LEP/non-LEP status.

In addition to paying particular attention to the achievement of LEP students in the general outcomes analysis, achievement data for LEP students who might be lost in English versions of standardized tests should be captured. Spanish language test data should be analyzed for schools in four large districts, as noted above. This analysis should assess whether the relationship between CSR and achievement gains is the same for LEP students tested in English and those tested in Spanish.

Finally, it is also important to examine the relationship of CSR to the reading and mathematics achievement for children with disabilities. Therefore, all achievement analyses should compare students with and without disabilities as a function of CSR. Analysis of the reading readiness data should also help determine the relationship between CSR and achievement for students with disabilities.

While reading achievement and readiness are core outcome variables in the evaluation plan, other important student outcomes may also be related to CSR. For example, the plan calls for a determination of whether referrals to special programs such as sheltered English courses, resource classes, and Reading Recovery and other reading programs have decreased. Changes in the number of students classified as special education, the promotion rates from third to fourth-grade, and the number of students referred for disciplinary action should also be investigated. The same analytic strategies outlined above for examining the relationship between CSR and reading achievement should be used in examining the relationships between CSR and these variables.

Qualitative Data Analyses. The qualitative data analyses deserve special discussion. Interviews and field notes should be entered into a computer-based program for organizing text data. The coding method should be valid, reliable and consistent across field workers.

To illustrate relationships derived from the analysis of qualitative data, visual displays, such as matrices and narrative tables, should be developed. Cross-site analyses should also be conducted which draw conclusions using both quantitative survey and qualitative interview data. For example, student achievement data should be triangulated with quantitative data on teacher characteristics and qualitative data on the nature of professional development to further provide insight into potential relationships among variables.

Advisory Board

An advisory board to represent the concerns of all interested parties, to maintain the independence of the study, and to help interpret the results in ways that are meaningful to those groups is also an integral part of the evaluation plan. The board should help frame the study, establish priorities among research questions, review instruments, interpret results and disseminate findings. The board should include representatives from:

- State government;
- Professional education organizations such as:
  - California School Boards' Association;
  - Association of California School Administrators;
  - California Federation of Teachers, United Teachers of L.A.; and
  - California Parent Teacher Association;
- Foundations;
- Research and academic organizations such as:
  - UCLA Center for Research on Evaluation, Standards, and Student Testing (CRESST);
  - Commission on Future of Teaching and Learning;
  - California State University Institute for Education Reform; and
  - California Center for School Restructuring.

Deliverables

A minimum of four reports should be written as part of the evaluation. The first three reports are designed to report on the formative part of the evaluation. The first is due not later than December 31, 1998, the second not later than December 31, 1999, the third due not later than December 31, 2000. These reports should report on analyses of data for each of the three years of the evaluation during which data are collected. Each of these annual formative reports should be structured around the seven major issues that comprise the Conceptual Model, and should contain data analyses designed to answer the set of research questions associated with each. Student outcome data (the seventh issue in the Conceptual model) are of particular importance. Analyses and reports should reflect this emphasis on student outcomes. Finally, each of the reports should contain a set of recommendations for how California's CSR program can be improved with respect to cost-efficiency and equity.

The fourth and final report should be summative and report on the overall findings from the evaluation. It, too, should be organized around the seven major issues described above. It should be completed by June 30, 2001.

References

Angrist, J. D. and Lavy, V. (1997). *Using Maimonides' rule to estimate the effect of class size on scholastic achievement.* Unpublished manuscript.

Blatchford, P. and Mortimore, P. (1994). The issue of class size for young children in school: What can we learn from research? *Oxford Review of Education*, 20(4), 411-428.

Blattner, B., Hall, K., and Reinhard, R. (1997). *Facilities and class size reduction.* Sacramento: School Services of California.

Burstein, L., McDonnell, L. M., Van Winkle, J., Ormseth, T., Mirocha, J., and Guiton, G. (1995). *Validating National Curriculum Indicators.* MR-658-NSF Santa Monica: RAND.

Cahen, L. S., Filby, N., McCutcheon, G., and Kyle, D. W. (1983). *Class size and instruction.* New York: Longman.

Comer, J. (1984). Home-school relationships as they affect the academic success of children. *Urban Society*, 16, 323-337.

Educational Research Services. (1980). *Class size reduction: A critique of recent meta-analyses.* Arlington, VA: Educational Research Services.

Finn, J. D. and Achilles, C. M. (1990). Answers and questions about class size: A statewide experiment. *American Educational Research Journal*, 27(3), 557-577.

Finn, J. and Voelkl, K. (1992). *Class size: An overview of research*, Department of Counseling and Educational Psychology Occasional Paper No. 92-1. Buffalo, NJ: State University of New York, Graduate School of Education Publications.

Glass, G., Cahen, L., Smith, M., and Filby, N. (1982). *School class size: Research and policy.* Beverly Hills: Sage Publications.

Illig, D. (1997). *Early implementation of the class size reduction initiative.* California: California Research Bureau.

Illig, D. (1996). *Reducing class size: A review of the literature and options for consideration.* California: California Research Bureau.

Klimes-Dougan, B., Lopez, J., Nelson, P., and Adelman, H. (1992). Two studies of low-income parents' involvement in schooling. *The Urban Review*, 24, 185-202.

Krueger, A. (1997). *Experimental estimates of education production functions*, draft paper, Princeton University.

Linn, R. L., and Kiplinger, V. L. (1995). *Linking statewide tests to the National Assessment of Educational Progress: Final report to the National Center for Education Statistics.*

Mitchell, D. E. and Beach, S. A. (1990). *How changing class size affects classrooms and students.* (Policy Brief). San Francisco: Far West Laboratory for Educational Research and Development.

Mosteller, F. (1995, Summer/Fall). The Tennessee study of class size in the early school grades. *Future of Children*, 5(2), 113-127.

Muller, C. (1993). Parent involvement and academic achievement: An analysis of family resources available to the child, in B. Schneider and J. Coleman (Eds.), *Parents, their children, and schools*. San Francisco: Westview Press, 77-113.

Murnane, R. J. and Levy, F. (1996, September 11). Why money matters sometimes. *Education Week*, 48, 36-37.

National Center for Education Statistics (NCES) (1992). *A profile of American eighth-grade mathematics and science instruction*. Technical report No. NCES-92-486. Washington, DC: U.S. Government Printing Office.

Office of Legislative Analysis (1996). *Class size reduction*. Author.

Pinnell, G.S., Pikulski, J.J., Wixson, K.K., Campbell, J.R., Gough, P.B., and A.S. Beatty (1995). *Listening to Children Read Aloud*. National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Shapson, S., Wright, E., Eason, G. and Fitzgerald, J. (1980). An experimental study of the effects of class size. *American Educational Research Journal*, 17(2), 141-152.

Stasz, C., Ramsey, K., Eden, R., DaVanzo, J., Farris, H., and Lewis, M. (1993). *Classrooms that Work: Teaching Generic Skills in Academic and Vocational Settings*. MR-169-NCRVE/UCB. Santa Monica: RAND.

Stevenson, D. and Baker, D. (1987). The family-school relation and the child's school performance. *Child Development*, 58, 1348-1357.

Stigler, J. and Fernandez, C. (1995). *Videotape classroom study: Field test report*. Los Angeles: University of California. (Available from author).

Wagner, B. and Phillips, D. (1992). Beyond beliefs: Parent and child behaviors and children's perceived academic competence. *Child Development*, 63, 1380-1391.

Yin, R. K. (1994). *Case study research: Design and methods*. Thousand Oaks, CA: Sage Publications.

Zykowski, J. (1996). *Student grouping: A direct link to achievement - A review of CERC's research on class size*. Riverside, CA: California Educational Research Cooperative (CERC).

[10/20/97]

The Importance of Evaluating Class Size Reduction to California Education Policy

There is growing enthusiasm for smaller class size as a means to boost student performance and to affect a host of related improvements in public education (e.g., reduce disciplinary problems, improve teacher morale, increase the opportunity for instructional innovation). Twenty-seven states, as well as the federal government, are either considering or implementing class size reduction programs. However, none of the state programs is as ambitious as is California's Class Size Reduction (CSR) initiative. In terms of expenditures, number of students served, and the cultural and economic diversity of the population it will address, the California initiative dwarfs those of any other single state. For these reasons, the California program has drawn the attention of state and federal planners as a crucial test case for CSR.

Proponents of California's CSR program point to a number of research studies which indicate that smaller class size improves student performance. Among the most prominent of these evaluations are Tennessee's STAR program, a recent ETS study using a sample of 177 districts and, even more recently, the SAGE study in Wisconsin. All three of these studies found that smaller class size improved student performance. In addition, "at-risk" students, i.e., those from minority or low income backgrounds, were particularly likely to benefit from placement in smaller classes. This result is especially significant for California public schools, which serve a large number of at-risk students. Because California has had one of the highest teacher/student ratios in the country, as well as some of the lowest rates of student achievement, CSR appears to many as a promising, and straight-forward, response to an important social problem.

However, there are reasons to question whether this initiative, as currently configured, will live up to expectations. In fact, the particular manner in which CSR is being introduced in California may lead to unexpected, and potentially negative, outcomes. Further, some of these outcomes may have deleterious effects on the populations that smaller class size is otherwise most likely to benefit.

Both the potential promise and drawbacks of the CSR program recommend a thorough, well conducted, evaluation. If the program does appear to improve student achievement, an evaluation can shed light on the conditions under which the program appears to be most efficacious as well as where it appears not to work at all. If no gains are observed in student achievement an evaluation will help policy makers understand why the program failed to work. In both cases, the findings can be the basis for generating recommendations for possible changes in the program.

Policy deliberations regarding the continuation, expansion, or curtailment of CSR will benefit greatly from reliable and timely information from an evaluation. For this reason, an intensive, multi-year evaluation of CSR is imperative. A recent report conducted by the California Research Bureau, under the auspices of the Legislative Analyst's Office (LAO), states that "...it ultimately will be necessary that the state conduct a comprehensive evaluation of the CSR initiative to clearly understand its impact."

Among the important policy issues to be examined in our proposed evaluation are:

- *How Responsive is the CSR Funding Formula to District Needs?*

  California's CSR initiative provides uniform funding for implementation, irrespective of local costs. In FY-97 this amounted to $650 per student in classes reduced to 20 or fewer students. In most districts these funds did not meet the cost of implementing the initiative. The FY-98 allocation has been increased to $800 per student, which is more than sufficient for many districts, but may still be inadequate to meet the needs of the poorest districts. In contrast, President Clinton's proposal for class-size reduction would follow the Title 1 allocation procedure which would result in poor, large districts receiving more funding than more affluent districts.

  Paradoxically, the California funding system may have provided a supplement for some wealthy districts, while placing a substantial burden on more financially strapped districts. A report by School Services of California (SCC) Inc. (April 18, 1997), estimated that of the 173,062 children included in the CSR, nearly one-third (31.7%) were in relatively affluent districts where the marginal costs of implementing CSR were less than the state grant they received for participating in the CSR program. At the other extreme, 13.7 percent of the children were enrolled in districts reporting marginal costs substantially greater than the state grant (i.e., an average costs of $1,000 to $2,000 per child in relation to the state grant of $650).

  Currently, districts can only claim CSR funds for classrooms they successfully convert to 20-to-1. Some school districts have larger classes to begin with, and therefore faced higher costs in achieving a 20-to-1 ratio of students to teachers. In addition, although some school districts had excess classroom space going into this reform, others had to procure portables (which have now become scarce in the state), and, particularly in inner-city areas, some schools districts are even short of spare land on which to place portables. It is likely that fewer classrooms in poorer districts will be able to meet the 20-to-1 goal (therefore producing fewer funds overall), and the local costs of meeting this goal will be much higher.

- *To What Degree Has CSR Displaced Other Important Programs Or Resources?*

  In efforts to participate in the CSR program, districts facing high local costs to implement this program (i.e., poorer districts) may have to draw the supplemental funds needed from elsewhere in their district and school budgets. In addition to costs, CSR may detract precious time and energy from school and district officials, and thereby interrupt the implementation of other reforms. This means that reduced class size might have to come at the expense of supplemental and support programs (e.g., libraries, counselors, social workers, nurses, child care); elective programs (e.g., art, music, and computers); space for support and supplemental programs (e.g., computer labs, language labs, band rooms, libraries, teachers' rooms), and perhaps higher enrollments in the non-CSR grades. These programs not only support teaching and learning but may, in fact, be vital for improving student achievement. A study conducted by the Joint Legislative Audit Committee in the first year of CSR suggests

that such displacements were common among districts participating in the state's CSR program.

- *How has CSR Implementation Affected Resource Allocation to Middle and High Schools?*

  As districts make trade-off decisions in order to implement CSR, they may need to take resources from their middle and high schools. In some cases, districts are moving grade 6 to middle schools in order to free up classroom space in elementary schools. Even in districts that do not need to transfer resources away from upper grades to due to CSR, these grades are not sharing in the substantial new resources now available for state education, overall.

  California was near the bottom among all states in regards to middle-school and high school funding before CSR was initiated. Due to CSR, the state might fall even farther behind in terms of the levels of resources made available for the education of post-elementary students. Inadequate resources at the middle and high school levels can lead to decreases in student achievement, negating the hoped-for positive effects of the small class sizes in the primary grades. Additionally, the recent TIMSS study found that the academic standing of American students compared to their counterparts in other countries declines at the junior and senior high school levels. As cited in the report "Our high school students are performing below international averages in math and science".

- *How has CSR Affected the Quality of the California Teaching Force?*

  With the implementation of the CSR initiative, the demand for teachers statewide far exceeds the supply. Over 18,000 new teachers were added to the statewide system this past year, of which almost one-quarter are working under an emergency credential or waiver. Another estimated 9800 new teachers were added this year, of which roughly one-third are estimated to be working with an emergency credential or waiver. The overall quality of workforce has gone down and with potentially negative impact on achievement. Teacher hiring and migration resulting from this initiative is likely to disproportionately affect the poorer districts. These districts often have more difficult working conditions and lower salaries. Consequently, they will find it even harder to find new teachers needed as a result of implementing CSR. Additionally, some of the credentialled teachers they had prior to this initiative (and, perhaps the best of these teachers) are likely to have considerable opportunities to move to higher paying neighboring districts. CSR may also cause collective bargaining problems for some districts. A report issued by the Legislative Analysts Office (LAO), found that nearly half the districts surveyed expect that CSR will complicate their negotiations with teachers. They expect teachers to raise concerns that the cost of CSR will lead to lowered salary increases. Districts may also contend with perceived inequities in the teaching burdens borne by teachers in CSR versus non-CSR classrooms.

- *In What Ways Has CSR Affected The Facilities Crisis In Public Schools?*

  Reducing student/teacher ratios creates an increased demand for classroom space. In their efforts to meet this demand, districts may take space away from school-site health clinics, parent training programs, and child care facilities. With the added pressure of welfare-to-work on poor neighborhoods, the loss of public school space for these support programs may negatively impact teaching and learning. Impoverished districts may be disproportionately affected by the costs of converting available space to classrooms, and may more often resort to sacrificing community support activities.

- *What Directions Should CSR Take in the Future?*

  Currently CSR is limited to grades k-3. However, some policymakers are already contemplating an expansion of the initiative. Before doing so, it may be appropriate to consider the various ways in which the program might be extended, and the relative merits and costs of these changes. For example, should class size reduction be targeted only to the lower grades or should higher grades be reduces as well? Should CSR resources be focused mainly on disadvantaged populations, or should they extend to all public school children? Should the ceiling for reduced classes be lower than 20? What is a pedagogically optimal, yet practicable, student/teacher ratio? Should expenditures be restricted to teachers with certain minimal qualifications as opposed to those on emergency credentials? Although our evaluation will not yield all the information needed to answer these questions, it will supply data upon which will add perspective to them.

The Stakes of CSR Success or Failure

The multiple burdens of funding, space, staffing, and loss or curtailment of other programs that the CSR initiative might impose raises the following question: does the state's particular implementation of CSR create problems which reduce the net benefits that smaller class size might otherwise afford? This question may be especially relevant in regards to at-risk populations. These populations tend to be those most responsive to smaller class size, yet they are served by the districts most likely to be disrupted by the California CSR initiative.

For the program to sustain the public support that the initiative has thus far enjoyed, it may have to produce sizable net gains in achievement. Failure in this regard may have significant repercussions for public education. The enthusiasm that has been expressed for the program represents a notable turn-around in public support of public education in California. Not only is there wide-spread backing for the expense that this program presents, but there is preliminary evidence that parents are willing to send their children to public rather than private schools. However, if the CSR program does not deliver its promised benefits, public support may not only erode for class size reduction, but for other investments in public education as well.

Comparing the Proposed Study with the Tennessee STAR Study

Proponents of the California CSR initiative place considerable weight on the Tennessee STAR study, in which the benefits of reduced class size were demonstrated in the context of a controlled experiment. However, there are important distinctions to be made between the circumstances that characterized the STAR program, and those that obtain in California.

Most students in the STAR study were white or African American, while nearly one-third of California's K-3 student population are not native English speakers. Tennessee had no shortage of fully credentialed teachers. In its first year, California's CSR implementation required hiring 18,400 new teachers. Of these new teachers, half were inexperienced, 30 percent were uncredentialed, and 21 percent were hired on emergency permits--meaning they had college degrees and had passed a competency test, but had no formal preparation for teaching. Tennessee's schools had sufficient classroom space to accommodate their class size-reduction effort. In its first year of implementation, California's schools needed 18,000 new classrooms. This meant that libraries, music rooms, computer labs, childcare centers, faculty lounges, and stages in auditoriums were converted into primary classrooms. Finally, Tennessee's schools had a year's planning time to implement CSR while California's schools had only a few months for the first year's implementation. For all these reasons, the STAR program may not be an infallible forecast of CSR in California.

The Importance of an Evaluation for California Education Policies

The various issues surrounding CSR suggest that an accurate understanding of the program's impact cannot be measured by annual reviews of student test scores, alone. The program, unlike nearly any other educational reform, directly affects nearly every element of public education in the state. Knowing how these elements are each affected by CSR, the repercussions these effects have on other elements, and the cumulative outcome all these effects have on student performance, is the best way to fully gauge the program's impact. Lacking this information, it will be difficult for policy makers (as well as administrators, teachers and parents) to understand why the program succeeds or fails to succeed, and why its effects may be differentially expressed across different regions and among different populations. Without such a complete understanding, further decisions regarding this costly and far-ranging reform may be compromised. In addition, the lack of periodic reports deprives districts of the information they need to make incremental, "on-line", corrections to their local CSR implementations.

There is also an important distinction to be made between the CSR program's perceived and actual effects. Performance gains associated with CSR may, in fact, be attributable to extraneous factors such as the implementation of the new STAR test, leading to spurious gains as teachers "teach to the test". Alternatively, the absence of performance gains during the initial implementation of CSR may be due to relatively transitory impediments that CSR creates, such as the integration of new teachers or the location of additional space. Without a proper evaluation, which would take such variables into account, the CSR program may be misrepresented as a success or as a failure. Either type of misinterpretation is likely to affect the nature of subsequent policy. For example, spurious gains may lead to inappropriate expansion of a program of questionable merit, while obscured success may cause the dismantling of one that does considerable good.